

WHAT IS CLAIMED IS:

1. A method of updating speech models for speech recognition,
comprising the steps of:
 - 5 identifying speech data for a predetermined set of utterances from a class of users, said utterances differing from a predetermined set of stored speech models by at least a predetermined amount;
 - collecting said identified speech data for similar utterances from said class of users;
 - correcting said predetermined set of stored speech models as a function of the
 - 10 collected speech data so that the corrected speech models are an improved match to said utterances than said predetermined set of stored speech models; and
 - updating said predetermined set of speech models with said corrected speech models for subsequent speech recognition of utterances from said class of users.
- 15 2. The method of claim 1 wherein said step of identifying speech data comprises comparing said utterances to the stored sets of speech models, obtaining a best match between the utterance of a user and a stored speech model in said predetermined set, and identifying as said speech data the utterance that differs from the best matched speech model by at least said predetermined amount.
- 20 3. The method of claim 1 wherein said step of collecting comprises saving identified utterances from said class of users, and saving correction data for the saved utterances representing corrections needed to minimize those differences between the respective utterances and said best matched speech models.
- 25 4. The method of claim 3 wherein a class of users is determined by registering users in accordance with predetermined criteria that characterize the speech of said class.

5. The method of claim 4 wherein said predetermined criteria include the primary language spoken by said user, the gender of said user, the age of said user the weight of the said user, the height of the said user, and the number of years the user has spoken the language of said utterances.

5

6. The method of claim 4 wherein said predetermined criteria include samples of calibrated utterances of the user.

7. The method of claim 1 wherein said predetermined set of stored
10 speech models are corrected as a function of the saved supplementary data when the number of said saved identified utterances exceeds a predetermined threshold.

8. The method of claim 7 wherein said step of updating comprises storing
15 in a centralized data base the corrected predetermined set of stored speech models, training new speech models in accordance with said corrected speech models, and distributing from said centralized data base to individual user sites said trained speech models

9. A method of building speech models for recognizing speech of users
20 of a particular class, comprising the steps of:

registering users in accordance with predetermined criteria that
characterize the speech of said particular class of users;

collecting a set of registration utterances from a user;

determining a best match of each said utterance to a stored speech
25 model;

collecting utterances from users of said particular class that differ from
said stored, best match speech model by at least a predetermined amount; and

retraining said stored speech model to reduce to less than said
predetermined amount, the difference between the retrained speech model and said
30 identified utterances from said users of said particular class.

10. The method of claim 9 wherein said predetermined criteria includes an identification of the primary language spoken by a user.

11. The method of claim 9 wherein said predetermined criteria includes an
5 identification of the gender of a user.

12. The method of claim 9 wherein said predetermined criteria includes an
identification of the number of years a user has spoken the
system target language.

10

13. The method of claim 9 wherein said predetermined criteria includes an
identification of the age of the user.

15

14. The method of claim 9 wherein said predetermined criteria includes an
identification of the height of the user.

15. The method of claim 9 wherein said predetermined criteria includes an
identification of the weight of the user.

20

16. The method of claim 9 wherein users register by transmitting to a
central data base information representing the primary language spoken by gender,
height, weight and age of a user, number of years the system target language has been
spoken by a user, age when the system target language was learned by a user and
samples of calibrated speech of a user.

25

17. The method of claim 9 wherein an utterance is sensed by sampling
speech of a user, and extracting from the sampled speech identifiable speech features.

18. The method of claim 9 wherein speech models of said identifiable
30 speech features are stored, and wherein a stored model of speech features that best
matches the extracted speech features is determined.

19. The method of claim 9 wherein the step of retraining includes transmitting to a central data base said collected features and correction data.

20. The method of claim 19 wherein the step of retraining further includes
5 using said collected features and to build new speech models.

21. The method of claim 20 wherein the step of retraining further includes returning said new speech models from said central data base to relevant user terminals, using at said user terminals both said new speech models and said stored
10 speech models to determine respective best matches of new utterances of users, and replacing at said user terminals the stored speech models with said new speech models after a predetermined number of utterances are determined to be better matched to said new speech models than to said stored speech models.

15 22. The method of claim 9 wherein said stored speech model is a hidden Markov model, but could be adapted to other classification schemes, e.g., dynamic time warping.

23. A method of creating speech models for speech recognition,
20 comprising the steps of:
 registering users in accordance with predetermined criteria that characterize the speech of a particular class of users;
 generating digital representations of utterances from said users;
 collecting from said particular class of users those digital
25 representations of similar utterances that differ by at least a predetermined amount from a set of stored speech models that are determined to be a best match to said utterances, and collecting corrections to said set of stored speech models that reduce the differences between an utterance and said set of models to a minimum;
 building a set of updated speech models based on said collected
30 corrections when the number of utterances that differ from said stored best match set of speech models by at least said predetermined amount, exceeds a threshold; and

using said set of updated speech models as said stored set of speech models for further speech recognition.

- 5 24. A system for updating speech models for speech recognition, comprising:
 plural user processors each programmed to:
 identify acoustic subword data for a predetermined set of
 utterances from a class of users, said utterances differing from a
 predetermined set of stored speech models by at least a predetermined
 amount;
10 collect said identified acoustic subword data for similar
 utterances from said class of users; and
 correct said predetermined set of stored speech models as a
 function of the collected acoustic subword data so that the corrected speech
 models are a closer match to said utterances than said predetermined set of
15 stored speech models; and
 a central processor, programmed to update said predetermined set of speech models at
 user processors with said corrected speech models for subsequent speech recognition
 of utterances from said class of users.
- 20 25. The method of claim 21 wherein a user processor is programmed to
 identify acoustic subword data by comparing said utterances to the stored sets of
 speech models, obtaining a best match between the utterance of a user and a stored
 speech model in said predetermined set, and identifying as said acoustic subword data
 that the utterance that differs from the best matched speech model by at least said
25 predetermined amount.
26. The method of claim 21 wherein a user processor is programmed to
 collect by saving identified utterances from said class of users.

27. The method of claim 23 wherein a class of users is determined by registering users in accordance with predetermined criteria that characterize the speech of said class.

5 28. The system of claim 24 wherein said predetermined criteria is selected from the primary language spoken by said user, the gender of said user, the age of said user, the weight of said user, the height of said user, the number of years said user has spoken the language of the utterances, and the age at which the target language is learned.

10

29. The system of claim 24 wherein said predetermined criteria include samples of calibrated utterances of the user.

15 30. The system of claim 24 wherein a user processor is programmed to correct said predetermined set of stored speech models as a function of the saved correction data when the number of said saved identified utterances exceeds a predetermined threshold.

20 31. The method of claim 27 wherein said central processor stores in a centralized data base the corrected predetermined set of stored speech models, and is programmed to train new speech models in accordance with said corrected speech models, and to distribute from said centralized data base to individual user processors said trained speech models.

25 32. A system for building speech models for recognizing speech of users of a particular class, comprising:

plural user processors, each programmed to:
sense an utterance from a user;

determine a best match of said utterance to a stored speech model; and
collect data from users of said particular class utterance that differ from said stored
best match speech model by at least a predetermined amount;

a central processor programmed and coupled to the plural processes

5 for:

registering users in accordance with predetermined criteria that
characterize the speech of said particular class of users; and

retraining said speech model stored at a user processor to
reduce to less than said predetermined amount the difference between the retrained
10 speech model and said identified utterances from said users of said particular class.

33. The system of claim 29 wherein said predetermined criteria includes
an identification of the primary language spoken by a user.

15 34. The system of claim 29 wherein said predetermined criteria includes
an identification of the gender of a user.

35. The system of claim 31 wherein said predetermined criteria includes
an identification of the number of years a user has spoken the system target language.

20

36. The system of claim 29 wherein said central processor is programmed
to register users by receiving from said user processors class information representing
the primary language spoken by gender, age, height , weight, number of years the
system target language has been spoken by, age when the system target language is

25 learned and samples of calibrated speech of respective users.

37. The system of claim 29 wherein a user processor is programmed to
sense an utterance by sampling speech of a user, and extracting from the sampled
speech identifiable speech features.

30

38. The system of claim 34 wherein the user processor stores speech models of said identifiable speech features, and is programmed to determine the stored model of a speech feature that best matches an extracted speech feature.

5 39. The system of claim 35 wherein the user processor is further programmed to produce correction data for extracted speech features which would reduce differences between said extracted speech features and the best matched stored models to less than a predetermined threshold.

10 40. The system of claim 36 wherein the user processor is programmed to collect those features extracted from utterances of a user which differ by at least said predetermined amount from the best matched stored models, together with the correction data for those extracted speech features.

15 41. The system of claim 37 wherein the user processor is programmed to transmit to said central processor said collected features and correction data for use at said central processor to retrain said speech models.

20 42. The system of claim 38 wherein the central processor is programmed to retrain said speech models by using said collected features and correction data to build new speech models that differ from said collected features by less than said predetermined threshold.

25 43. The system of claim 39 wherein the central processor is programmed to return said new speech models to said user processors, and said user processors are further programmed to use both said new speech models and said stored speech models to determine respective best matches of new utterances of users and to replace the stored speech models with said new speech models after a predetermined number of utterances are determined to be better matched to said new speech models than to
30 said stored speech models.

44. The system of claim 29 wherein said stored speech model is a hidden Markov model.